DOCUMENT RESUME

ED 307 278 TM 013 200

AUTHOR Vancouver, Jeffrey B.; And Others
TITLE Determination of Validity and Bias in the Use of GPA and MCAT in the Selection of Medical School Students.
PUB DATE Mar 89
NOTE 16p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Admission Criteria; Ethnic Groups; *Grade Point Average; Higher Education; Medical Schools; *Medical Students; Minority Groups; Multiple Regression Analysis; *Predictive Validity; *Selective Admission; *Test Bias; Test Validity
IDENTIFIERS Majority Groups; *Medical College Admission Test; National Board of Medical Examiners

ABSTRACT

The ability of the Medical College Admissions Test (MCAT) and undergraduate grade point average (GPA) to predict success in medical school was studied, and two complementary methods of determining if the tests are biased against ethnic groups were examined. Data from 497 majority and 82 minority medical students at the College of Human Medicine at Michigan State University (East Lansing) over a 6-year period were used to test for predictive validity and differential prediction. Majority status was defined as all other groups except Black, Mexican American, Mainland Puerto Rican, and Native American. The National Board of Medical Examiners Part I (NBME-I) examination was used as a measure of student performance in medical school. Science GPA and a composite MCAT score (the biology, chemistry, physics, reading, and quantitative subtests) were evaluated as predictors. Moderated multiple regression and the Cleary model (Cleary, 1968) were used to determine whether test bias was present in science GPA or MCAT scores. The interaction of ethnicity with the predictors was also evaluated. Both the science GPA and the composite MCAT scores were valid and predictive of success in medical school as measured by the NBME-I. Both were equally valid for minority and majority groups. There were significant mean differences between the groups, but ethnicity did not affect the meaning of the scores in terms of predicting success on the NBME-I. Moderated multiple regression was the more sensitive measure of differential validity; the Cleary model can confirm results of a moderated multiple regression equation. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED307278

DETERMINATION OF VALIDITY AND BIAS IN THE USE OF GPA AND MCAT IN THE SELECTION OF MEDICAL SCHOOL STUDENTS

by

Jeffrey B. Vancouver, M.A., Mary Ann Reinhart, Ph.D.,
David J. Solomon, Ph.D., and James J. Haf, Ph.D.
Michigan State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

San Francisco

March 1989

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
JEFFREY B. VANCOUVER
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Send Correspondence To:

Jeffrey B. Vancouver
A212 E. Fee Hall
OMERAD
Michigan State University
East Lansing, MI 48824

Running Head: VALIDITY AND BIAS

TM013200

BEST COPY AVAILABLE

INTRODUCTION

A primary concern of administrators selecting students for admission into medical school is the potential for unfair discrimination of applicants from different ethnic groups. Two credentials frequently used to differentiate among student applicants are the Medical College Admissions Test (MCAT) and students' undergraduate grade point average (GPA). When the inferences derived from the use of MCAT scores and GPA have different ability to predict success in medical school for different groups, they are said to have differential prediction. For example, if high marks on the MCAT and GPA predict successful performance for one group, but the same marks predict poor performance or have no predictive ability for another ethnic group, then differential prediction exists. This source of bias has the potential to limit the opportunities for acceptance for one or all ethnic groups.

In the study reported here, we first examined the ability of MCAT scores and GPA to predict medical school performance and then examined two complimentary methods of determining if the tests are biased against ethnic groups. These methods are easy to use and are recommended for all medical colleges who are concerned about potentially biasing the opportunities of ethnic groups.

The use of different measures to select medical school applicants has been examined by Mitchell (1987) and Jones and Mitchell (1986). In a survey of North American medical schools, Mitchell (1987) found that admissions personnel rated MCAT and

under- graduate GPA high in importance for medical student selection. Jones and Mitchell (1986) examined the predictive validity and the differential prediction of MCAT scores on a dichotomous measure of academic difficulty. Academic difficulty was defined as either delayed graduation or withdrawal/dismissal from medical school. The authors found predictive validity for MCAT scores, that is, lower MCAT scores predicted academic difficulty. However, MCAT scores had differential prediction between minority and majority students. Even with similar MCAT scores, minority students were more likely to have academic difficulties than did their white counterparts. Jones and Mitchell did not investigate the predictive validity or differential prediction of GPA.

In their study, Jones and Mitchell (1986) used a dichotomous measure of success in medical school. One potentially more powerful measure of early success in medical school is the National Board of Medical Examiners Part I (NBME-I) examination. The three parts of the boards are administered during subsequent phases in a student's basic science and clinical education. Part I is frequently administered after the second year in medical school, and it can be seen as a comprehensive examination of mastery of basic medical science information. NBME-I is a continuous measure of level of achievement, permitting the use of more powerful methods of detecting bias in the predictors than were used in the Jones and Mitchell study.

In the study reported here, we first tested the predictive validity of GPA and MCAT scores and secondly, used two complimentary methods to test for differential prediction in the use of GPA and MCAT scores. NBME-I was used as a measure of medical student performance in all analyses.

METHODOLOGY

Data from 497 majority and 82 minority medical students attending the College of Human Medicine at Michigan State University over a six year period were used to test for predictive validity and differential prediction. Minority status was defined as Black, Mexican American, Mainland Puerto Rican, and Native American. Majority status was defined as all other ethnic groups. Scores on NBME-I were used to evaluate student performance at the end of the first two years of medical school.

Selection Instruments. Two predictors of NBME-I scores (GPA and MCAT) were selected based on their use by most medical colleges and on our correlational analyses of a number of potential credentials. Science GPA (S-GPA) and overall GPA were both moderatedly correlated with NBME-I scores (.33 and .32, respectively). In the interest of parsimony, we chose to use S-GPA as one predictor of NBME-I scores.

A composite MCAT score was formed by averaging the biology, chemistry, physics, reading and quantitative subtest scores. The subtests correlated between .37 and .53 with NBME-I. A confirmatory factor analysis (Hunter & Gerbing, 1979) conducted on the MCAT subtests indicated that the subtests formed a single

factor, which correlated .57 with NBME-I. We used the composite MCAT score as the second predictor of NBME-I scores.

Methods Used to Test for Differential Prediction. We used two methods to determine whether test bias was present in S-GPA and MCAT scores: moderated multiple regression (Bartlett, Bobko, Mosier, & Hannan, 1978) and the Cleary model (Cleary, 1968). Our first method for determining test bias was moderated multiple regression. In this analysis S-GPA and MCAT were entered into an regression analysis to predict NBME-I. We examined the results to determine whether both S-GPA and MCAT predicted NBME-I scores when they were simultaneously entered into the regression analysis (predictive validity). If the change in the multiple R squared associated with the predictors was significant, the predicitve validity of both predictors would be established and both would be retained. If the change in the multiple R squared was not significant, redundancy in information would be established and only MCAT would be retained due to its higher initial correlation with NBME-I.

Next we created a dichotomous variable of ethnic status of the student (minority, majority), which was entered into the regression equation following S-GPA and MCAT. We then examined the results of the analysis to determine if ethnicity predicted NBME-I scores by examining the change in the multple R squared.

In the last step in this procedure, we examined the interaction of ethnicity with each of the predictors (S-GPA, MCAT). A statistically significant interaction indicates that the

regression lines calculated for the two ethnic groups are not parallel. This means that the ability of a predictor to estimate NBME-I differs depending on the ethnicity of the student. A finding of a significant interaction of ethnicity with S-GPA is differential prediction for S-GPA. A significant interaction of ethnicity with MCAT indicates differential prediction for MCAT.

The second method we used to assess test bias was the Cleary model (1968) of regression analysis. In this model, a regression line based on S-GPA and MCAT scores was computed for each ethnic group. Then, both the regression weights and the constants of the regression lines for each group were compared to determine if they were significantly different. A finding of statistically different regression weights or constants is an indication of differential validity (Hulin, Drasgow, & Parsons, 1983).

RESULTS

We first used moderated multiple regression to search for differential prediction, followed by the Cleary model for confirmation. The results of the moderated multiple regression are presented in Table 1. S-GPA and MCAT were significantly correlated with NBME-I. The multiple R squared of S-GPA and MCAT with NBME-I was .355, $p = .001$. The addition of the ethnicity factor did not significantly change the multiple R squared (change in R squared = .0027, p = NS). More importantly, the interaction terms of ethnicity with S-GPA and ethnicity with MCAT did not significantly change R squared (change in R squared = .0009, p = NS).

To further test the findings of predictive validity but no differential validity, we also used the Cleary model on the data. In this analysis, multiple regression was used to produced two regression equations. Table 2 shows the regression weights and the constants with the standard errors in parentheses. We tested for significant differences between the two S-GPA weights, the two MCAT weights, and the two constants. None of the differences was significant, confirming the moderated multiple regression results showing no differential prediction.

Because no evidence of differential prediction was found, we were able to calculate a regression equation based on the total sample. The larger sample size of the total sample provides the most stable regression equation. Calculations based on this equation could then be used to predict applicants' future NBME-I scores. A useful means of doing this is described by Solomon, Vancouver, Reinhart, and Haf (1989), who used the regression equation to create nomograms. The regression equation calculated from all student data is provided in Table 2.

DISCUSSION

The findings indicate that using S-GPA and a composite MCAT score based on the biology, chemistry, physics, reading and quantitative subtests is valid and equally predictive for minority and majority groups. Contrary to the findings of Jones and Mitchell (1986), no differential prediction was found between minority and majority groups. Significant mean differences in S-GPA and MCAT were found between the groups, but the meaning of

the scores in terms of predicting success on NBME-I was not affected by ethnicity. Thus, similar scores between the groups on S-GPA and MCAT are equally predictive of NBME-I performance.

This finding is not surprising given the lack of finding differential prediction in many cognitive based selection instruments (Hunter, Schmidt, & Hunter, 1979). Hunter et al, (1979) argued that, in fact, accumulations of studies, which allow for the correction of many statistical artifacts, demonstrate that differential prediction probably does not exist when cognitive tests are used as selection instruments. Our findings reinforce these conclusions for medical colleges by showing a fairness in using S-GPA and MCAT to predict students' mastery of comprehension of basic science information as measured by NBME-I.

CONCLUSION AND IMPLICATIONS

In the interest of fairness to all ethnic groups, selection variables should be assessed for predictive validity of medical school success and all measures should be free of differential prediction. Since moderator multiple regression is the most sensitive measure of differential validity, medical colleges can use it to assess the fairness of selection measures. The Cleary model can provide additional information about where the differential information exists, and should also be used. If differential validity does not exist, the Cleary Model can confirm the results of the the moderator regression equation.

Most colleges use GPA and MCAT to select medical students and both appear to be valid and fair measures of medical school success.

Although this finding is contrary to Jones and Mitchell's results, our analyses, using a more powerful measure of medical student mastery, are consistent with previous research investigating the results of many studies. We believe medical colleges can feel more confident using GPA and MCAT than previously thought.

Table 1

Regression Equation Values Derived from the Moderated Regression Analysis

Term	Multiple R	R^2	Change in R^2	p
S-GPA[1] & MCAT	.5957	.3548	.3548	.001
Ethnicity[2]	.5980	.3576	.0027	NS
Ethnicity X GPA & Ethnicity X MCAT	.5987	.3585	.0009	NS

[1]Science Grade Point Average

[2]Minority/Majority Status

Table 2

Regression Equation Values Derived for the Cleary Model Analyses

	Constant	Science GPA	Composite MCAT
Majority Sample	71.33	37.55	28.69
(N = 497)	(39.12)	(9.60)	(2.52)
Minority Sample	121.37	23.43	25.09
(N = 82)	(68.06)	(21.69)	(5.94)
Total Sample	51.22	39.87	29.86
(N = 579)	(27.70)	(8.19)	(2.02)

Note. Regression weights (and standard errors) are presented in the table.

REFERENCES

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.

Bartlett, C. J., Bobko, P. Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderatted multiple regression strategy: An alternative to differential analysis. Personnel Psychology, 31, 233-241.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory. Homewood, IL: Dow Jones - Irwin.

Hunter, J. E., & Gerbing, D. (1979). Unidimensional Measurement and Confirmatory Factor Analysis. Occasional Paper #20, Institute for Research on Teaching, Michigan State University.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A review and analysis. Psychological Bulletin, 86, 721-735.

Jones, R. F., & Mitchell, K. (1986, April). Racial/ethnic differences in the predictive validity of MCAT scores. Paper presented at the meeting of the American Education Research Association, San Francisco, CA.

Mitchell, K. J. (1987). Use of MCAT data in selecting students for admission to medical school. Journal of Medical Education. 62, 871-879.

Solomon, D. J., Vancouver, J. B., Reinhart, M. A., & Haf, J. J. (1989, January). A graphic method for presenting expectancy of passing NBME Part I examinations based on average preclinical anatomy and MCAT scores. Paper presented at the Winter Conference for the Michigan Educational Research Association, East Lansing, Michigan.

DETERMINATION OF VALIDITY AND BIAS IN THE USE OF GPA AND MCAT IN THE SELECTION OF MEDICAL SCHOOL STUDENTS

Two credentials frequently used to differentiate among student applicants are the Medical College Admissions Test (MCAT) and students' undergraduate grade point average (GPA). In the study reported here, we first examined the ability of MCAT scores and GPA to predict medical school performance and then examined two complimentary methods of determining if the tests are biased against ethnic groups. These methods are easy to use and are recommended for all medical colleges who are concerned about potentially biasing the opportunities of ethnic groups.

Data from 579 medical students attending a midwestern medical college over a six year period were used to test for predictive validity and differential prediction. We used two methods to determine whether S-GPA and MCAT scores predicted NBME Part I scores similarily for minority and majority students. The methods were moderated multiple regression (Bartlett, Bobko, Mosier, & Hannan, 1978) and the Cleary model (Cleary, 1968).

The findings indicated that using S-GPA and a composite MCAT score based on the biology, chemistry, physics, reading and quantitative

subtests is valid and equally predictive for minority and majority groups. These results contradict the findings of Jones and Mitchell (1986), which indicated differential prediction for ethnic groups of MCAT on academic difficulty in medical college. However, they are not surprising given the lack of findings showing differential prediction in many cognitive based selection instruments (Hunter, Schmidt, & Hunter, 1979). We believe medical colleges can feel more confident using GPA and MCAT than previously thought.

Keywords: Bias, Selection, MCAT, GPA, NBME Part I